

ETIČKI IZAZOVI UMJETNE INTELIGENCIJE I ROBOTIKE

Sažetak: Rad obrađuje više etičkih izazova umjetne inteligencije i robotike. Nakon uvodnih napomena o etičkim aspektima inženjerstva i kratkog prikaza triju osnovnih teorija normativne etike, kao implicitan izazov razmatra se prijetnja slobodi volje, smislenosti etike i ljudskoj posebnosti. Središnji dio rada zauzima razmatranje sedam eksplicitnih izazova: autonomna vozila, autonomni naoružani sustavi, socijalna robotika, upotreba umjetne inteligencije i robota u medicini, prediktivna analitika, utjecaj umjetne inteligencije i robotike na ljudska zaposlenja te primjena novih tehnologija za razne vrste ljudskog poboljšanja. U zaključku se daju neke smjernice za bavljenje etičkim pitanjima umjetne inteligencije i robotike, ali se upozorava i na njihov mogući utjecaj na samo razumijevanje etike i bavljenje njome.

Ključne riječi: Umjetna inteligencija, robotika, etika.

Uvod

Razvoj i sve raširenija primjena umjetne inteligencije i robotike u različitim područjima ljudskog života i djelovanja nose sa sobom različite etičke izazove. Nakon uvodnog razmatranja odnosa etike i inženjerstva, kao i kratkoga prikaza triju osnovnih teorija normativne etike (utilitarizma, deontologije i etike vrline), u radu se razmatra implicitne i eksplicitne etičke izazove ovih tehnologija. Ključni implicitni izazov sastoji se u potencijalu umjetne inteligencije i robotike da dovedu u pitanje slobodu volje, a time i samu smislenost etike i ukorijenjeno vjerovanje u ljudsku posebnost. Potom slijedi analiza sedam eksplicitnih izazova umjetne inteligencije i robotike: (1) autonomna vozila, (2) autonomni naoružani sustavi, (3) socijalna robotika, (4) upotreba umjetne inteligencije i robota u medicini, (5) prediktivna analitika, (6) utjecaj umjetne inteligencije i robotike na ljudska zaposlenja i (7) primjena novih tehnologija za različite vrste ljudskog poboljšanja. Zaključak rada sadrži neke preporuke o poželjnome načinu etičkog razmatranja ovih novih tehnologija, ali i o mogućem utjecaju tih istih tehnologija na ljudsko razumijevanje morala i samo bavljenje etikom.

1. Inženjerstvo i etika

Umjetna inteligencija i robotika sve više prožimaju brojne aspekte ljudskoga života, poput prometa, medicine, komunikacije, zabave, znanstvenih istraživanja, obrazovanja, radnih uvjeta, čak i osobnih i intimnih odnosa. Interakcija ljudi i ovih tehnologija nosi sa sobom specifične etičke izazove koji zaokupljaju pozornost ne samo filozofa i društvenih znanstvenika, nego jednako tako prirodnih znanstvenika i inženjera. Inženjerstvo, naime, postaje sve češćom temom etičkih rasprava, ali raste i interes samih inženjera za etičke aspekte njihova djelovanja. Posebnost inženjerstva, kao što ističu van de Poel i Royakkers (2011: 1), sastoji se upravo u tome što se ono “ne svodi tek na bolje razumijevanje svijeta, nego i na njegovo mijenjanje”, uslijed čega inženjerstvo – prvenstveno zahvaljujući vjerovanju samih inženjera da njihova

tehnološka rješenja i inovacije svijet čine boljim – predstavlja “inherentno moralno motiviranu djelatnost” koja iziskuje “etičku refleksiju i znanje”.

Strukovna udruženja inženjera također posvećuju sve veću pozornost etičkim aspektima njihova djelovanja, osobito na području umjetne inteligencije i robotike. Institut inženjera elektrotehnike i elektronike (IEEE), kao najveća i najutjecajnija svjetska organizacija inženjera, objavio je opsežnu studiju, *Ethically Aligned Design* (2019), s etičkim smjernicama za razvoj novih, na umjetnoj inteligenciji i robotici, zasnovanih tehnologija. Raste i broj stručnih izvješća, studija i deklaracija brojnih drugih tijela i institucija u kojima se upozorava na nužnost etičke regulacije umjetne inteligencije i robotike. Primjerice, UNESCO-ova Svjetska komisija za etiku znanstvenog znanja i tehnologije objavila je opsežno izvješće o etici robotike (COMEST 2017) i preliminarnu studiju o etici umjetne inteligencije (COMEST 2019), dok je Europska komisija osnovala stručnu skupinu koja priprema etičke smjernice za “pouzdanu” (*trustworthy*) umjetnu inteligenciju. Dakako, problem svih sličnih studija i izvješća je njihova općenitost i brzo zastarijevanje u odnosu na razvoj novih tehnologija i njihovih primjena. Stoga mnogi tehnički studiji u svoje nastavne programe sve više uvode kolegije u kojima se buduće inženjere poučava osnovama etike i etike tehnologije koje će, tijekom svoje karijere, moći primjenjivati i na one tehnologije koje će se tek pojaviti (tako se, primjerice, od akademske godine 2018./2019. na Fakultetu elektrotehnike i računarstva Sveučilišta u Zagrebu izvodi kolegij “Etika i nove tehnologije”).

Etička refleksija o tehnologijama poput umjetne inteligencije i robotike u pravilu se oslanja na širok raspon teorija i pojmova razvijenih u okvirima standardne filozofske etike (dakako, pritom se koristi i spoznajama drugih disciplina, poput sociologije ili prava). Radi lakše ilustracije konkretnih etičkih dilema koje se javljaju u pogledu ovih tehnologija, do kraja odsječka bit će ukratko naznačene osnovne crte triju utjecajnih etičkih teorija koje se pritom najčešće spominju: utilitarizam, deontologija i etika vrline (za raspravu o etici kao disciplini na hrvatskom vidjeti npr. Talanga 1999, Berčić 2012, Bracanović 2018).

1.1. Utilitarizam

Utilitaristi smatraju da je procjena posljedica ključna za donošenje suda o moralnoj ispravnosti ili pogrešnosti nekog djelovanja. Primjerice, ako će, u danim okolnostima, laganje imati bolje posljedice od govorenja istine, onda je naša moralna dužnost lagati; ako će pak govorenje istine imati bolje posljedice, onda je naša moralna dužnost govoriti istinu. Dakako, oko pitanja što točno jesu “dobre posljedice” unutar tradicije utilitarizma postoje neslaganja: za Jeremyja Bentham (1907 [1789]) to je “ugoda”, za Johna Stuarta Milla (1998 [1863]) “sreća”, za Petera Singera (2003 [1979]) “zadovoljene preferencije”. Ako se možda i ne slažu u pogledu definicije dobrih posljedica i “korisnosti” (*utility*) kao njihova temeljnog pojma, među utilitaristima postoji visoka razina slaganja oko drugih središnjih elemenata njihove teorije. Jedan takav element je načelo nepristranosti. Prema ovome načelu, moja vlastita “korisnost”, “ugoda” ili “sreća” ni po čemu nije važnija od “korisnosti”, “ugode” ili “sreće” bilo koje druge osobe. Stoga, prilikom odlučivanja o tome kako ću postupiti ne smijem pokazivati nikakvu pristranost prema samome sebi, već djelovati tako da, kako je tvrdio Bentham, stvorim “najveću sreću za najveći broj ljudi.” Ukratko: utilitarizam je teorija koja inzistira na nepristranom maksimiranju dobrih i minimiziranju loših posljedica u svijetu.

1.2. Deontologija

Deontolozi, za razliku od utilitarista, smatraju da posljedice djelovanja ne mogu biti mjerilo moralnosti djelovanja – kao što je smatrao Immanuel Kant (2016 [1785]) – ili da ne mogu biti jedino mjerilo njegove moralnosti – kao što je smatrao David Ross (2002 [1930]). Posljedice ne mogu biti mjerilo moralnosti jer to ima protuintuitivne i zdravorazumski neprihvatljive implikacije (primjerice, kada loše namjere slučajno urode dobrim posljedicama ili kada dobre namjere urode lošim posljedicama). Stoga deontolozi tvrde da se postupke mora procjenjivati s obzirom na namjere ili, konkretnije, s obzirom na “načela”, “pravila” ili “maksime” kojima se vode oni koji ih izvode. Kant je smatrao da načela koja stoje iza naših postupaka jesu moralno prihvatljiva ako ih možemo zamisliti kao univerzalne zakone (što je čuvena Kantova provjera moralnosti poznatija kao “kategorički imperativ”) i ako ne dovode do korištenja osoba kao pukih sredstava ili instrumenata (što je zahtjev da se prema drugima ponašamo poštujući njihova moralna prava i dostojanstvo osobe). Za razliku od utilitarizma koji je iznimno fleksibilna moralna teorija i dopušta da jedan te isti postupak, ovisno o svojim posljedicama, u nekim situacijama može biti ispravan, a u drugima pogrešan, deontologija je apsolutistička moralna teorija upravo zato što smatra da su neki postupci, poput ubijanja ili laganja, uvijek i u svakoj situaciji moralno pogrešni.

1.3. Etika vrline

Etika vrline – za koju se koriste još nazivi “kreposna etika” i “aretaička etika” – razlikuje se i od utilitarizma i od deontologije po tome što ne nastoji toliko odgovoriti na pitanje “Koji su postupci moralno ispravni, a koji moralno pogrešni?”, koliko na pitanja poput “Što je vrlo ili kreposna osoba?”, “Što je sretan život?” i “Koja je važnost vrline za sreću?”. Stoga se obično ističe da su utilitarizam i deontologija teorije koje su usredotočene na postupke (*act-centred theories*), a da je etika vrline teorija koja je usredotočena na djelatnika (*agent-centred theory*) odnosno na vrline i poroke kao svojstva djelatnikova karaktera. Povijesno najutjecajnija verzija teorije vrline jest ona Aristotelova (1992). Za Aristotela, etičke vrline kao što su pravednost, hrabrost ili velikodušnost stječu se navikavanjem i odgojem, a nužne su za ostvarivanje osobujno ljudske svrhe i sreće (*eudaimonia*), koja se postiže u okviru razumom vođenog života u društvenoj zajednici. Etika vrline – uslijed svog naglaska na postizanju ljudske sreće ili dobra – povezuje moral s mnogim drugim, često i izvanmoralnim, vrijednostima kao što su cjelovit i dobrima ispunjen život i odnosi s drugim ljudima. Većina zastupnika etike vrline u 20. stoljeću – kao što su Elisabeth Anscombe, Alasdair MacIntyre, Rosalind Hursthouse, Martha Nussbaum i Susan Wolf – u osnovi upozorava da etičko mišljenje ne može biti stvar mehaničke procjene pojedinačnih postupaka, kao što to pretpostavlja većina drugih etičkih teorija. Etika, kao učenje o ljudskome dobru, više je poput složene navigacije prema sreći i dobrom životu.

2. Implicitni izazovi

Predodžbe o umjetno stvorenim i inteligentnim bićima odavno su prisutne u ljudskoj povijesti. Većina pregleda razvoja umjetne inteligencije i robotike (npr. Perkowitz 2004) tako spominju mitološke priče (poput one o Hefestu i njegovim od zlata izrađenim sluškinjama ili one o bakrenom divu Talosu koji je branio Kretu), književna djela (poput romana Mary Shelley iz 1818. o “biću” koje je napravio dr. Frankenstein ili drame Karela Čapeka *Rossumovi*

univerzalni roboti iz 1920. u kojoj je riječ “robot” prvi put upotrijebljena) i znanstveno-fantastične filmove kao što su *Metropolis* (1927), *2001: Odiseja u svemiru* (1968), *Ratovi zvijezda* (1977) ili *Bladerunner* (1982). Većina ovih fikcionalnih prikaza umjetnih i inteligentnih bića nosi određene etičke poruke, a obično je to upozorenje na opasnosti koje njihovo stvaranje ili upotreba (ljudsko “igranje Boga”) može donijeti čovječanstvu. No ovim mitološkim, književnim ili znanstveno-fantastičnim prikazima opasnosti umjetnih bića zacijelo ne treba pridavati preveliku etičku težinu jer oni nisu ni zamišljeni kao argumentirane i na znanstvenim činjenicama utemeljene rasprave.

Sredinom 20. stoljeća, zahvaljujući prije svega napretku računalne znanosti, umjetna inteligencija izlazi iz nekadašnjih fikcionalnih ili imaginarnih okvira, a “uređaj sposoban za izvođenje funkcija koje se normalno povezuju s ljudskom inteligencijom, poput zaključivanja, učenja i vlastitog poboljšavanja” (Rosenberg 1986: 10) počinje se razmatrati kao ozbiljna teorijska i praktična mogućnost. Sam termin “umjetna inteligencija” (*artificial intelligence*) prvi put se pojavljuje 1955. u prijedlogu skupine znanstvenika da se na Dartmouth koledžu u Hanoveru (New Hampshire, SAD) organizira ljetni projekt posvećen njenom proučavanju. Znanstvenici koji su stajali iza ovog prijedloga i projekta – John McCarthy, Marvin Minsky, Nathaniel Rochester i Claude Shannon – vjerovali su da predloženo proučavanje može poći “od pretpostavke da se svaki aspekt učenja ili bilo koje drugo svojstvo inteligencije u načelu može opisati toliko precizno da se može načiniti stroj koji će ga simulirati” (McCarthy et al. 2006 [1955]: 12).

Istraživanja iz umjetne inteligencije danas su dobrim dijelom pragmatično orijentirana te uključuju ogranke kao što su procesiranje prirodnog jezika, reprezentacija znanja, automatizirano zaključivanje, strojno učenje, duboko učenje, računalni vid i robotika (Russell i Norvig 2016: 2-3). Iako ovi ogranci bilježe važne rezultate koji su omogućili stvaranje inteligentnih sustava za obavljanje najraznovrsnijih zadataka, većina stručnjaka se slaže da smo još uvijek daleko od stvaranja onoga što se naziva “opća” (*general*), “široka” (*wide*) ili “jaka” (*strong*) umjetna inteligencija: umjetna inteligencija koja bi imala istu razinu općenitosti i širinu primjene poput ljudske inteligencije i koja ne bi isključivo – što je to slučaj sa sustavima “uske” (*narrow*) ili “slabe” (*weak*) umjetne inteligencije – isključivo oponašala ljudsku inteligenciju obavljajući samo jedan zadatak ili vrlo ograničeni raspon zadataka. Drugim riječima, ako i neki sustav umjetne inteligencije može nadmašiti ljudsku inteligenciju u obavljanju jednog zadatka u ograničenom području (primjerice, upravljanju zrakoplovom na autopilotu), on se još uvijek ne može mjeriti s ljudskom inteligencijom u obavljanju mnoštva drugih, po svojoj naravi vrlo različitih zadataka. Kao što ističu Nick Bostrom i Eliezer Yudkowsky (2014: 318), “računalo Deep Blue jest postalo svjetski prvak u šahu, ali osim toga ne može igrati čak ni ‘dame’, a kamoli voziti automobil ili doći do nekog znanstvenog otkrića.”

Jedno od ključnih pitanja povezanih s umjetnom inteligencijom glasi: Može li se uopće za neki stroj reći da je “inteligentan” – ili da bi mogao postati “inteligentan” – na isti način kao što su to ljudi? Dva klasična, međusobno suprotstavljena, odgovora na ovo pitanje u 20. stoljeću dali su Alan Turing (1950) i John Searle (1980). Dok je Turing smatrao da će s vremenom biti razvijen stroj koji će toliko uvjerljivo oponašati ljudsku inteligenciju da će biti nemoguće razlikovati ga od čovjeka (svoj prijedlog testa za provjeru inteligencije stroja, danas poznat kao “Turingov test”, izvorno je nazvao “Imitation Game”), Searle je nastojao dokazati (pomoću misaonog argumenta nazvanog “Kineska soba”) da za neki stroj, bez obzira koliko on bio sofisticiran, nikada neće biti smisleno reći da doista “misli”, da je “inteligentan” ili da ima “razumijevanje”, u prvome redu zato što tek mehanički slijedi procedure koje su mu zadane programom. Kao što je to slučaj s većinom sličnih rasprava, rasprava o općoj umjetnoj

inteligenciji se nastavlja i u njoj sudjeluju stručnjaci iz različitih područja kao što su računalna znanost, filozofija, psihologija, kognitivna znanost, neuroznanost, lingvistika i dr.

Implicitan ili neizravan etički izazov koji proizlazi iz same mogućnosti opće umjetne inteligencije sastoji se u sljedećem: Ako inteligencija iste općenitosti i širine kao ona ljudska jest moguća i načelno ostvariva u nekom fizičkom sustavu poput stroja, onda je moguće da i ljudska inteligencija – i ljudski um kao njezin nositelj – i sama predstavlja neku vrstu složenog fizičkog (biološkog) sustava ili stroja. Da iskoristimo spomenutu formulaciju McCarthyja i njegovih kolega sa samih početaka istraživanja umjetne inteligencije: Ako se “svaki aspekt učenja ili bilo koje drugo svojstvo inteligencije u načelu može opisati toliko precizno da se može načiniti stroj koji će ga simulirati”, onda ljudsko mišljenje – i ljudsko ponašanje koje je motivirano tim mišljenjem – nije slobodno nego je uzročno determinirano kao što je uzročno determiniran bilo koji stroj. Mogućnost umjetne opće inteligencije, dakle, zaprijetila bi slobodi ljudske volje, a samim time i smislenosti etike, koja počiva na pretpostavci da su ljudi slobodna bića kojima ima smisla propisivati što trebaju, a što ne trebaju činiti. Ako je ljudsko mišljenje i ponašanje determinirano poput funkcioniranja stroja, etika gubi svoj smisao. Ovu implikaciju mnogi smatraju odbojnom i neprihvatljivom, ne samo zato što poništava slobodu ljudske volje i smislenost etike, nego i zato što poništava ljudsku posebnost i različitost u odnosu na ostatak živog, a možda i neživog svijeta.

3. Eksplicitni izazovi

Osim naznačenih implicitnih ili neizravnih etičkih izazova, umjetna inteligencija i robotika donose i neke sasvim eksplicitne ili izravne izazove. Konkretno primjene ovih tehnologija u mnogim područjima ljudskog života i djelovanja, naime, otvaraju specifična moralna pitanja i probleme. U nastavku slijedi razmatranje sedam takvih eksplicitnih etičkih izazova:

3.1. Autonomna vozila

Autonomna vozila (ponekad se koriste i nazivi “samovozeći automobili” ili “robotska vozila”) zauzimaju posebno mjesto u raspravama o etici umjetne inteligencije i robotike. Jedan od ključnih razloga za to je praktične naravi: Brojne svjetske tvrtke ubrzano rade na razvijanju i testiranju takvih vozila i ona će, prema mnogim procjenama, uskoro postati prometna stvarnost. Očekuje se da će uvođenje autonomnih vozila dovesti do znatnog smanjenja broja prometnih nesreća i pogibija na cestama (čiji najčešći uzrok je ljudska pogreška, poput vožnje u alkoholiziranom stanju, dekoncentracije ili umora), optimiziranja prometa i smanjenja prometnih gužvi, smanjene potrošnje goriva i zagađenja okoliša, lakšeg sudjelovanja u prometu za tjelesne invalide i starije osobe i sl.

Središnja etička rasprava u vezi s autonomnim vozilima (korisni pregledi su Millar 2017 i Nyholm 2018) vodi se oko njihovih “etičkih postavki” (*ethics settings*), odnosno pitanja kao što su: Kako bi autonomna vozila trebala reagirati u nepredviđenim situacijama u kojima će se suočiti s izborom između veće ili manje štete ili zla? Kako bi trebala odlučiti – točnije: kako bi ih trebalo programirati da odlučuju – prilikom izbora između usmrćivanja manjeg ili većeg broja pješaka ili prilikom izbora između usmrćivanja manjeg broja putnika u vozilu ili većeg broja pješaka? S ovim je pitanjima povezano i možda temeljnije pitanje o tome tko bi uopće trebao odlučivati kakve će biti “etičke postavke” autonomnih vozila: proizvođači vozila, njihovi korisnici ili država?

Utilitaristički odgovor na prva dva pitanja, u skladu s načelom “najveće sreće za najveći broj ljudi”, zacijelo bi glasio da autonomna vozila trebaju biti programirana tako da uvijek spase što je moguće više, odnosno da žrtvuju što je moguće manje ljudskih života, čak i ako to zahtijeva žrtvovanje putnika u samome vozilu. No nije isključeno da bi autonomna vozila koja bi uvijek žrtvovala manji broj ljudi (npr. pješake i vozače u susjednim vozilima) mogla postati stalnim izvorom straha i tjeskobe kod mnogih ljudi i stoga, zapravo, utilitaristički nepoželjna. Deontološki odgovor nedvojbeno bi bio drukčiji, poput nekih smjernica iz izvješća *Automated and Connected Driving* (2017) Etičkog povjerenstva Ministarstva prometa i digitalne infrastrukture Savezne Republike Njemačke. Autonomno vozilo, ističe se u izvješću, mora nastojati minimizirati štetu prilikom eventualnih prometnih nezgoda, ali ono ne bi smjelo biti programirano da bira između ljudskih života. Deontološki razlog iza ovih smjernica je vjerovanje da bi prepuštanje autonomnome vozilu odluke o tome tko će biti žrtvovan kako bi netko drugi bio spašen predstavljalo narušavanje ljudske autonomije, prava na život i dostojanstva, odnosno postupanje s osobama kao s pukim sredstvima, oruđima ili stvarima.

Pitanje tko bi uopće trebao odlučivati o tome kakve će biti etičke postavke autonomnih vozila potiče dileme koje, osim etičkih, uključuju pravne, ekonomske i socijalno-političke aspekte. Ako bi odluka bila na tvrtkama koje proizvode autonomna vozila, to bi donekle bilo konzistentno s uobičajenom zakonskom odgovornošću za sigurnost i ispravnost proizvoda, ali bi moglo značiti i preveliku odgovornost tvrtki za eventualne štete i ljudske žrtve koje bi takva vozila prouzročila (na što većina tvrtki, iz financijskih razloga, zacijelo ne bi pristala). Ako bi o etičkim postavkama odlučivali individualni vlasnici ili korisnici vozila, to bi vjerojatno rezultiralo time da bi većina automobila imala “egoistične” postavke (postavke koje uvijek spašavaju vozača ili putnika u vozilu na štetu ostalih sudionika u prometu), što bi dovelo do povećanja ukupnog broja žrtava u prometu. Ako bi pak o etičkim postavkama autonomnih vozila odlučivao zakonodavac ili država (primjerice, propisujući obvezne utilitarističke etičke postavke, kao što predlažu Gogoll i Müller 2016), to bi se moglo shvatiti kao prekomjerno uplitanje države u individualnu ljudsku slobodu i život (da ne spominjemo da većina ljudi vjerojatno ne bi htjela kupiti autonomno vozilo koje je programirano da ih žrtvuje u određenim situacijama).

3.2. Autonomni naoružani sustavi

Suvremeno ratovanje uvelike je obilježeno upotrebom naprednih tehnoloških sustava poput daljinski upravljanih naoružanih dronova i krstarećih projektila. Sudeći po ulaganjima vojno-industrijskih kompleksa mnogih zemalja u tehnologije poput umjetne inteligencije i robotike, izvjesno je da će ratovanje budućnosti biti dodatno obilježeno upotrebom poluautonomnih ili posve autonomnih naoružanih sustava: sustava koji će sami na bojišnici “odlučivati” o tome na koji će način i protiv koga primijeniti neku vrstu smrtonosne sile. Rasprava o takvim sustavima redovito izaziva negativne reakcije, u prvome redu uslijed raširenog uvjerenja da je rat toliko nemoralna pojava da je besmisleno, možda i licemjerno, raspravljati o mogućnosti njegova moralnog vođenja. Slična logika vodi do zaključka da etički prihvatljivo korištenje autonomnih naoružanih sustava, uslijed činjenice da su oni namijenjeni ubijanju i ranjavanju, jednostavno nije moguća. Ratovi predstavljaju ljudsku stvarnost i mnogi od njih bili su posebno poznati po svojoj brutalnosti i velikom broju poginulih i ranjenih vojnika i civila. No ipak, ponavljanje takvih ratova nastoji se spriječiti na razne načine, između ostaloga i donošenjem međunarodno obvezujućih propisa i pravila (kao što su Ženevske konvencije i Haaške konvencije) o tome na koji se način rat smije, a na koji način ne smije voditi. Mnogi autori, u tome smislu, smatraju

da i razmatranja o etički opravdanoj ili neopravdanoj upotrebi autonomnih naoružanih sustava ipak nisu *a priori* besmislena (korisne rasprave su npr. Krishnan 2009, Galliot 2015, Strawser 2013).

Rasprave o moralnoj ili nemoralnoj upotrebi autonomnih naoružanih sustava najčešće se odvijaju u okviru teorije pravednog rata, odnosno u okviru njezinih dvaju ogranaka: (1) *jus ad bellum*, kao raspravi o uvjetima moralno opravdanog pribjegavanja ratu, te (2) *jus in bello*, kao raspravi o tome koja sredstva vođenja rata jesu moralno dopuštena, a koja nisu (koristan pregled je Primorac 2006). Dva *jus in bello* uvjeta posebno su relevantna za procjenu moralnosti autonomnih naoružanih sustava: zabrana ubijanja neboraca (npr. civila i neborbenog vojnog osoblja) te zabrana izazivanja štete ili zla koje bi bilo nerazmjerno ostvarenom vojnom cilju. Dobar dio autora zauzima negativan stav prema autonomnim naoružanim sustavima. Noel Sharkey (2012) smatra, primjerice, da bi sustavi u kojima nema čovjeka “u petlji” (*in the loop*) vjerojatno doveli do učestalijeg kršenja zabrane ubijanja neboraca jer takvi sustavi, uslijed svojih tehničkih ograničenja, ne bi bili u stanju razlikovati vojnike od civila ili vojnike koji se bore od vojnika koji su ranjeni ili se žele predati. Robert Sparrow (2007) problematičnim smatra to što, u slučaju da takvi sustavi počine ratni zločin, ne bi bilo moguće jasno utvrditi tko je za nj odgovoran. Mogući kandidati između kojih bi odgovornost vjerojatno bila podijeljena i tako raspršena su: časnik koji je zapovjedio upotrebu sustava, njegovi dizajneri ili programeri, ministarstvo obrane koje je naručilo izradu ili kupovinu sustava, a možda predsjednik države kao vrhovni vojni zapovjednik.

Drugi autori razvijanje i upotrebu naoružanih autonomnih sustava promatraju u nešto pozitivnijem svjetlu, pod uvjetom, dakako, da se radi o sustavima koji su tehnički pouzdani i za čije je funkcioniranje jasno na kojemu članu ljudskog zapovjednog lanca leži odgovornost. Gert-Jan Lokhorst i Jeroen van den Hoven (2012) smatraju da bi takvi sustavi – zahvaljujući svojim naprednim tehničkim svojstvima, poput preciznosti, brzine ili dobrih senzora – mogli značajno smanjiti broj i vojnih i civilnih žrtava u ratnim sukobima. Pretpostavka pritom glasi da upravo “neljudskost” takvim sustavima daje određene prednosti pred ljudima koji sudjeluju u vojnim operacijama, u smislu da neće ubijati nasumično jer na njih, kao strojeve, ne mogu utjecati emocije poput straha, mržnje ili želje za osvetom. Autonomni naoružani sustavi mogli bi biti programirani i da ubijanje ograniče na najmanju nužnu mjeru (ili da ubijanje zamijene ranjavanjem). Čak i ako njihovo pridržavanje etike ratovanja i ne bi bilo savršeno, moguće je da bi napredak bio postignut već i time što bi bili bolji od ljudi.

3.3. Socijalna robotika

Socijalna robotika bavi se dizajniranjem humanoidnih i nehumanoidnih robota čija je primjena raznolika, poput robota za druženje, edukacijskih robota, terapijskih robota, robota za skrb o starijim osobama i djeci ili robota za intimne i seksualne odnose. Korisne strane ovih robota nije teško prepoznati jer oni su upravo namijenjeni tome da, primjerice, olakšaju i produlje samostalan život starijim ili bolesnim osobama, pomažu u nastavi s djecom, obavljaju poslove u bolnicama ili ustanovama za skrb za koje je teško pronaći radnu snagu, pomažu u terapiji osoba s određenim mentalnim poteškoćama, omogućuju neku vrstu intimnog ili seksualnog zadovoljstva osobama koje takvo što teško ostvaruju (poput tjelesnih invalida) ili jednostavno obavljaju određene zadaće u domaćinstvu i tako oslobađaju ljudima vrijeme za druge životno važne aktivnosti.

Budući da u pravilu neće izazvati neku izravnu ili očitu štetu, poput gubitka ljudskih života (kao što je to slučaj s autonomnim vozilima i autonomnim naoružanim sustavima), etičke

izazove socijalnih robota nije jednostavno formulirati, pogotovo ne iz perspektive etičkih teorija usredotočenih na posebne postupke kao što su deontologija ili utilitarizam. Mnogi autori (npr. Turkle 2012, Sharkey i Sharkey 2012b, Scheutz 2012) ipak se slažu oko sljedeće okvirne procjene: Prepuštenost trajnoj interakciji sa socijalnim robotima može dovesti do slabljenja interakcije s ljudima i izazvati neku vrstu emocionalne ovisnosti. Štoviše, budući da će korisnici takvih robota biti starije osobe, djeca ili osobe s psihičkim poteškoćama, pojavit će se i opasnost obmane: vjerovanja da se nalaze u interakciji s drugom osobom, a ne robotom. Prekomjerna izloženost djece socijalnim robotima mogla bi imati negativan utjecaj na razvoj njihovih socijalnih vještina i vrline, poput tolerancije, za koje je potrebna interakcija s drugim osobama i posebice vršnjacima. Uslijed ovih opasnosti, često upozorenje u mnogim etičkim kodeksima robotičara glasi da robote ne treba “antropomorfizirati” ili činiti “humanoidnijima” preko mjere koja je neophodna za njihovo učinkovito obavljanje funkcija za koje su predviđeni. Etička teorija čije su zasade zacijelo u najvećem raskoraku sa svijetom socijalne robotike je etika vrline. Socijalni roboti, naime, prijetu narušavanjem međusobne uvjetovanosti i sklada koji – prema zastupnicima etike vrline – postoji između ljudske sreće ili dobrobiti s jedne strane te društveno i emocionalno ispunjenog života s druge strane.

3.4. Umjetna inteligencija i robotika u medicini

Medicina je područje u kojemu su nove tehnologije oduvijek izazivale moralne dileme, kao što su respiratori, stroj za hemodijalizu, pacemaker, prenatalna dijagnostika, umjetna oplodnja ili razne tehnike transplantacije. Moralne dileme u osnovi nastaju zbog toga što medicina danas može izvesti sve više toga što u ne tako davnoj prošlosti nije bilo ni zamislivo. Primjerice, pacijente se danas može umjetnim sredstvima dugo održavati na životu, što potiče rasprave o moralnosti različitih vrsta eutanazije (poput dobrovoljne i nedobrovoljne te aktivne i pasivne), definiciji smrti (kardiopulmonalna smrt, smrt višeg mozga ili smrt moždanog debla) te dostupnosti i pravednoj distribuciji dostupnih organa za transplantaciju. Tehnike prenatalnog otkrivanja genetskih poremećaja zametka ili fetusa na sličan način potiču rasprave o moralnoj opravdanosti pobačaja.

Umjetna inteligencija i robotika, kao tehnologije koje se u medicini sve češće koriste, donose nove ili pojačavaju stare medicinske etičke izazove. Jedan izazov dolazi od robotske kirurgije: upotrebe robota koji su stanju sami ili pod većom ili manjom kontrolom ljudskog kirurga izvoditi složene operativne zahvate. S jedne strane, upotreba takvih robota predstavlja veliki napredak i poželjna je jer omogućuje preciznije, sigurnije i manje invazivne intervencije u ljudsko tijelo. Njome se izbjegava i uobičajene probleme koje izaziva “ljudski čimbenik” prilikom sličnih zahvata, poput umora, dekoncentracija ili drhtanja ruku. S druge pak strane, postoji opasnost primjene robotskih kirurških sustava bez njihova odgovarajućeg ili dostatnog testiranja, odnosno opasnost da se pacijente, kojima prijeti ozbiljna bolest ili smrt i koji su stoga pod velikim psihološkim pritiskom, brzopletu i nepromišljeno usmjerava na operacije pomoću takvih sustava, a da im se nije objasnilo potencijalne rizike i komplikacije (Sharkey i Sharkey 2012b). Slično kao i s autonomnim vozilima i autonomnim naoružanim sustavima, javlja se problem odgovornosti za robotske kirurške zahvate s kojima je nešto pošlo po zlu: snosi li odgovornost liječnik koji koristi robota, bolnica koja ga je kupila, njegov proizvođač ili njegovi dizajneri i programeri?

Slični problemi opterećuju i medicinske ekspertne sustave, čiji najpoznatiji primjer je Watson tvrtke IBM. Zahvaljujući umjetnoj inteligenciji i mogućnosti gotovo trenutnog pristupa golemoj količini medicinskih podataka – medicinskih udžbenika, znanstvenih članaka i

zdravstvenih kartona pojedinačnih pacijenata – koja je neizmjereno veća od količine podataka kojom raspolaže bilo koji ljudski liječnik, Watson može, koristeći prirodni jezik, postavljati i provjeravati dijagnoze i predlagati liječenja za pacijente. Sve i ako ostavimo po strani sumnje u stvarnu učinkovitost ovakvih sustava (npr. Müller 2018), postoji stanovita zabrinutost (npr. Lu 2016) da bi oni mogli dovesti do pretjeranog oslanjanja liječnika na tehnologiju, narušavanja njihove spremnosti da pamte stara i stječu nova medicinska znanja, te gubitka vještine fizičkog pregleda i suosjećajne komunikacije s pacijentima. Dodatne komplikacije mogle bi se pojaviti u vezi s praksom “informiranog pristanka”. Može li se pacijentu doista, kako bi bio informiran i na osnovi toga mogao pristati ili odbiti liječenje, objasniti na koji točan način algoritmi nekog medicinskog ekspertnog sustava dolaze do svojih dijagnoza i prijedloga za liječenje? Nema sumnje da će uvođenje novih medicinskih tehnologija utemeljenih na umjetnoj inteligenciji i robotici dovesti do novih moralnih dilema koje će zahtijevati određene izmjene postojećih kodeksa medicinske etike.

3.5. Prediktivna analitika

Etički osjetljive su i primjene umjetne inteligencije radi obrade velikih skupova podataka (*Big Data*), kao što je to slučaj s prediktivnom analitikom. Prediktivna analitika je vrsta analize podataka čiji primarni cilj je predviđanje budućih događaja ili trendova. Prediktivna analitika se koristi u mnogim područjima. Raširena je u “personaliziranom marketingu” – oglašavanju proizvoda ili usluga koje ne cilja na opću populaciju, nego je “personalizirano” ili prilagođeno individualnim korisnicima. Često spominjan primjer ove upotrebe prediktivne analitike je kampanja kojom je *Target*, trgovački lanac u Sjedinjenim Državama, na osnovi podataka prikupljenih od velikog broja stalnih kupaca, uspio s visokom preciznošću predvidjeti koje žene među njihovim kupcima su trudne. Ovo predviđanje omogućilo im je da trudnicama, prije nego što će one roditi, pošalju personalizirane oglase i kupone za dječje proizvode poput kolijevki ili pelena i na taj ih način pridobiju za svoje kupce. Kampanja je bila vrlo uspješna i donijela je *Targetu* milijune dolara (prema Duhigg 2012).

Nešto drukčiji primjer primjene prediktivne analitike potječe iz 2012. kada je pedesetak stručnjaka bilo angažirano u izbornom stožeru Baracka Obame. Zadatak im je bio ne samo identificirati neodlučne glasače nego i predvidjeti koja vrsta kontakta će pridobiti njihov glas: telefonski poziv, dolazak volontera na kućni prag, letak ili televizijski oglas? Očekivalo se, osim toga, da identificiraju i one glasače koje je najbolje “ostaviti na miru” jer će ionako glasati za Obamu, a neki oblik kontakta samo bi ih mogao uznemiriti ili potaknuti da promijene svoje glasačke preferencije. Znanje o ovakvim pojedinostima je vrijedno jer povećava učinkovitost političke kampanje, usmjeravajući i ljudske i financijske resurse prema glasačima koje je moguće pridobiti, odnosno usmjeravajući je od glasača koje kampanja neće pridobiti i od glasača koje nije potrebno pridobivati ili koje bi pokušaj pridobivanja mogao odbiti. Rašireno je mišljenje da je korištenje prediktivne analitike 2012. bio jedan od ključnih čimbenika prevage Baracka Obame pred njegovim protukandidatom Mittom Romneyem (prema Siegel 2013).

Od 2016. djelatnici pozivnog centra za prijavu zlostavljanja i zapostavljanja djece u okrugu Allegheny (Pennsylvanija, SAD) opremljeni su računalnim alatom AFST (*Allegheny Family Screening Tool*). Radi se o algoritmu koji daje “drugo mišljenje” o svakoj pojedinačnoj prijavi, analizirajući veliku količinu podataka o djetetu i članovima njegove obitelji pohranjenu u više različitih baza podataka (npr. o korisnicima socijalne pomoći, izdržavanju zatvorskih kazni, zloupotrebi droge ili alkohola i sl.). AFST ima svrhu pomoći djelatnicima centra da objektivnije procjenjuju zaprimljene pozive, posebice treba li poziv zanemariti kao neutemeljen

ili poslati djelatnike socijalne službe da hitno interveniraju u prijavljenu obitelj. Na osnovi analize više od 100 mogućih pokazatelja i kriterija, AFST za nekoliko sekundi daje procjenu razine opasnosti za svako dijete. Svrha mu je svratiti pozornost djelatnika na pozive koje bi olako zanemarili kao neutemeljene, ali i na pozive koje bi pogrešno procijenili kao alarmantne i poslali socijalnu službu u obitelji koje to nisu zaslužile. Prema službenim podacima dostupnim na mrežnim stranicama okruga Allegheny, ulaganja u AFST od oko milijun dolara isplatila su se.

Mnoge reakcije na prediktivnu analitiku su negativne. Jedna od standardnih kritika je da ona može narušiti pravo na privatnost, posebice kada velike tvrtke u tolikoj mjeri profiliraju svoje kupce da mogu vrlo precizno predvidjeti njihovo ponašanje i preferencije, pa čak i o njima zaključiti krajnje osobne i intimne stvari poput trudnoće (*Target* je dospio u središte medijske pozornosti kada je otac jedne tinejdžerice doznao za njezinu trudnoću pronašavši u poštanskom sandučiću njihove kupone). Javlja se i bojazan da bi metode poput prediktivne analitike dati preveliku moć državnim tijelima, političkim strankama ili interesnim skupinama da ciljano utječu na važne političke i društvene procese i time umanjuju politički utjecaj građana (Furnas 2012). Jednako tako, budući da algoritmi takvih računalnih alata mogu presudno utjecati na individualne živote, postavlja se pitanje njihove transparentnosti: ako je algoritam preporučio neki postupak na našu štetu, onda želimo znati kako je do te preporuke došlo i nije li algoritam programiran ili “treniran” na pristran način odnosno ne uključuje li neke ljudske predrasude. Ovo je posebice važno u primjeni takvih algoritama u donošenju sudskih odluka, odluka o odobravanju raznih zajmova i subvencija ili, kao što smo vidjeli, medicinskih dijagnoza ili odluka o intervencijama socijalnih službi.

3.6. Umjetna inteligencija, robotika i radna mjesta

Važno etičko i socijalno-političko pitanje u vezi s razvojem umjetne inteligencije i robotike jest ono o njihovom utjecaju na tržište rada, posebice na nestanak brojnih poslova koje su obavljali ljudi. Uslijed automatizacije i uvođenja robota u proizvodnju odavno su nestala brojna radna mjesta (automobilska industrija je paradigmatičan primjer) i sasvim je izvjesno da će još mnoga radna mjesta nestati u budućnosti. Pritom se ne radi tek o radnim mjestima koja se mogu relativno lako automatizirati, poput zavarivanja ili lakiranja u tvorničkim halama, nego i o radnim mjestima koja su u intelektualnom ili čak kreativnom smislu zahtjevnija, poput knjigovođa, službenika na šalterima ili prevoditelja. Kao što ističe Byron Reese (2018), uvođenje novih tehnologija u proizvodnju oduvijek je izazivalo društvene potrebe: kada je u 16. stoljeću William Lee izumio stroj za pletenje, primjerice, kraljica Elizabeta odbila je izdati mu patent vjerujući da će to njene podanike ostaviti bez posla i učiniti prosjacima; u 19. stoljeću pripadnici ludističkog pokreta uništavali su tvorničke strojeve u znak prosvjeda zbog ukidanja radnih mjesta uvođenjem novih tehnologija; londonski list *The Times* je 29. studenoga 1814. prvi put tiskan pomoću parnog tiskarskog stroja, a prosvjede i prijetnje radnika smirilo je tek obećanje vlasnika novina da će ih zadržati sve dok ne pronađu slične poslove negdje drugdje.

Općenito se vjeruje da će umjetna inteligencija i robotika imati dva učinka na tržište rada: (a) dio radnih mjesta će nestati jer će poslove predviđene tim radnim mjestima kvalitetnije i/ili jeftinije obavljati inteligentni strojevi; (b) zahvaljujući ovim novim tehnologijama, bit će stvorena nova radna mjesta koja ranije ili nisu postojala ili za kojima nije postojala ozbiljnija potreba. Jedan od problema ovakvog razvoja dugoročan je odnos između (a) i (b), ali i posljedice koje će gubitak radnih mjesta imati na pojedinačne radnike: Hoće li oni koji su ostali bez posla moći pronaći novi posao? Hoće li u tome uspjeti u nekom, za prosječan ljudski život,

razumnom roku? Hoće li se morati (i moći) prekvalificirati? Postoje li poslovi koji će preživjeti uvođenje robota i umjetne inteligencije? Odgovore na ovakva pitanja iznimno je teško dati jer oni pretpostavljaju predviđanje dugoročnih ekonomskih trendova (što je samo po sebi složeno i nesigurno) i predviđanje dugoročnih znanstvenih i tehnoloških trendova (što je možda još složenije i nesigurnije).

Odnos umjetne inteligencije, robotike i svijeta rada otvara i pitanja distributivne pravednosti. Radna mjesta su izvor prihoda i pristupa raznim dobrima (poput stanovanja ili obrazovanja). Ako nove tehnologije utječu na dostupnost ovih dobara velikom broju ljudi, možda bi njihov utjecaj morao biti reguliran u skladu s načelima pravednosti. Ali s kojim točno načelima? Neki će tvrditi, poput zastupnika *laissez-faire* ili libertarijanske teorije (Nozick 1974), da vlasnici tvrtki koje zarađuju zahvaljujući ulaganju u robotiku i umjetnu inteligenciju imaju slobodu činiti što god žele sa svojim vlasništvom i da nemaju obvezu zaposliti bilo koga tko im nije potreban. Vjerojatno bi dodali i da bi nekakva obveza zapošljavanja ljudskih radnika ili dodatno oporezivanje sputalo njihovu poduzetničku i inovatorsku inicijativu od koje čitavo društvo ima koristi. Alternativa bi bila neka vrsta socijalno-liberalnog shvaćanja pravednosti, poput onog koje je zastupao John Rawls (1999 [1971]). Prema ovome shvaćanju, društveno-ekonomske nejednakosti u društvu se mogu dopustiti, ali pod uvjetom da se time ne narušava temeljne građanske slobode, da svi članovi društva imaju jednake mogućnosti dolaska do socijalno-ekonomski probitačnijih dužnosti i položaja, te takve nejednakosti donose što je moguće veću korist najlošije stojećim članovima društva. Ovakvo poimanje pravednosti, prema nekim mišljenjima (npr. Sandbu 2017), iziskivalo bi ili posebne poreze, poput “poreza na robote” koji bi se koristio za pomoć ljudima koji su zbog njih izgubili poslove, ili uvođenje univerzalnog osobnog dohotka (*universal basic income*). Obje ideje su prijeporne i o njima se vode oštre rasprave.

3.7. Nove tehnologije i ljudska poboljšanja

Specifična rasprava povezana s mnogim novim tehnologijama usredotočena je na pitanje mogu li one radikalno utjecati na ljudsku prirodu i – ako mogu – trebamo li takvo što spriječiti ili poticati. Rasprava je započela neovisno o umjetnoj inteligenciji i robotici, kada su biomedicina, genetika i farmakologija omogućile sredstva i metode ne samo liječenja, nego i poboljšanja ljudi (*human enhancement*) preko granica koje su statistički ili za ljude kao biološku vrstu “normalne”. U literaturi se kao tri česte teme pojavljuju tjelesna poboljšanja (poput snage, brzine, imuniteta, vida ili sluha), kognitivna poboljšanja (poput pamćenja ili brzine zaključivanja) i moralno poboljšanje (poput suzbijanja agresivnosti, poticanja altruističnih sklonosti ili pojačavanja psihološke motivacije za moralno djelovanje). Metode kojima će se navodno moći postići ova poboljšanja uključivat će farmakološka sredstva (razne vrste lijekova poput Modafinila, Ritalina ili Prozaca), kirurške zahvate (poput plastične ili rekonstruktivne kirurgije), genetski inženjering (odabir poželjnih genetskih svojstava za potomstvo), ali i neke zahvate u kojima će umjetna inteligencija i robotika igrati važnu ulogu (neke vrste proteza i implantata, egzoskeleti, endoskeleti, bionički udovi i pužnice, sučelja mozak-računalo i sl.).

Protivnici poboljšanja, kao što je Michael Sandel (2009), smatraju da ona donose neke opasnosti koje nije moguće formulirati jednostavno i pomoću osnovnih etičkih pojmova kao što su “autonomija”, “pravednost” ili “ljudska prava”. Prema Sandelu, poboljšanja bi mogla dovesti gubitka divljenja prema bilo kojem individualnom ljudskom postignuću, trudu ili talentu, pretjeranog roditeljskog uplitanja u (genetsku) sudbinu svoje djece i nestanka solidarnosti. Zagovornici poboljšanja, kao što je John Harris (2007), smatraju da poboljšanja –

pod uvjetom da su njihovi rizici svedeni na razumnu mjeru – nisu moralno problematična i da predstavljaju logičan nastavak težnje ljudi da svoj život i živote svojih bližnjih učine što kvalitetnijim. Alberto Giubilini i Julian Savulescu (2018), kao zagovornici “moralnog poboljšanja”, smatraju da bi umjetna inteligencija mogla biti u službi ljudske moralnosti. Oni argumentiraju da bilo korisno stvoriti “umjetnog moralnog savjetnika” (*artificial moral advisor*): vrstu umjetne inteligencije koja bi ljudima pomagala da učinkovitije i konzistentnije donose moralne odluke u okolnostima kada ne raspolažu s dovoljno relevantnih informacija ili kada su emocionalno pristrani ili pod utjecajem predrasuda. No neki autori u prijedlozima moralnog poboljšanja vide opasnosti. Prema Nicholasu Agar (2012), moralno poboljšani ljudi mogli bi činiti posebnu podvrstu “transljudi” (*transhumans*) ili “transosoba” (*transpersons*) koje bi, upravo zbog svoje moralno poboljšane prirode, imale viši moralni status od “običnih” ljudi (možda na način da bi, primjerice, autonomna vozila u dilematičnim situacijama uvijek žrtvovala “obične” ljude). Ova ideja gotovo da nas vraća na početak razmatranja, pokazujući kako bi nekadašnji tek implicitni ili neizravni etički izazovi umjetne inteligencije i robotike – oni povezani s ljudskom slobodom, sposobnošću moralnog izbora i posebnosti – na jedan novi način mogli postati sasvim eksplicitni ili izravni.

Zaključak

Rasprave o etičkim izazovima umjetne inteligencije i robotike sve su razgranatije, u prvome redu uslijed razgranatosti njihovih primjena. Stoga je zacijelo besplodno voditi opću raspravu o njihovim etičkim aspektima i opasnostima jer svako posebno područje njihove primjene, prema svemu sudeći, otvara neke probleme koji se ne moraju nužno javljati u drugim područjima (etičke implikacije socijalne robotike, primjerice, zasigurno nemaju previše toga zajedničkog s etičkim implikacijama industrijskih robota ili autonomnih naoružanih sustava). Drugim riječima, raspravu o etici novih tehnologija poput umjetne inteligencije i robotike potrebno je kontekstualizirati i ne očekivati da će rješenja koja smatramo prihvatljivima u jednome području biti prihvatljiva i u drugim područjima. Osim toga, ne treba izgubiti iz vida mogućnost da će umjetna inteligencija i robotika utjecati i na način na koji ljudi poimaju etiku i pristupaju moralnim pitanjima. Nije isključeno da će ljudi, uslijed utjecaja ovih i s njima povezanih tehnologija, izgubiti dio svog moralnog senzibiliteta i neke probleme uopće prestati promatrati kao moralne probleme (narušavanje privatnosti je vjerojatno jedan od najizglednijih kandidata za takvo što). Možda će ljudski sustav moralnih vrijednosti, uslijed sve različitijih etičkih izazova koje u svakodnevni život unose sve različitije primjene moderne tehnologije, postati još više fragmentiran i nekonzistentan? Hoće li se odgovornost ljudi smanjiti ili će se, uslijed tehnološki omogućenog povećanja njihove moći, možda porasti? Osim implicitnih i eksplicitnih etičkih izazova razmotrenih u ovome radu, ovo su također neka pitanja kojima će se etika novih tehnologija u budućnosti vjerojatno morati baviti.

Bibliografija

Agar, N. 2013. "Why is it possible to enhance moral status and why doing so is wrong?", *Journal of Medical Ethics* 39: 67-74.

Aristotel, 1992. *Nikomahova etika*, prev. T. Ladan (Hrvatska sveučilišna naklada: Zagreb).

Automated and Connected Driving, 2017. Ethics Commision, Federal Ministry of Transport and Digital Infrastructure. Dostupno na: <https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.html?nn=355056> [stranica posjećena: 8. svibnja 2019.]

Bentham, J. 1907. [1789] *An Introduction to the Principles of Morals and Legislation* (Clarendon Press: Oxford).

Berčić, B. 2012. *Filozofija*, sv. 1 (Ibis grafika: Zagreb).

Bostrom, N. i Yudkowsky, E. 2014. "The ethics of artificial intelligence", u: K. Frankish i W. M. Ramsey (ur.), *The Cambridge Handbook of Artificial Intelligence* (Cambridge University Press: Cambridge), 316-334.

Bracanović, T. 2018. *Normativna etika* (Institut za filozofiju, Zagreb).

COMEST 2017. *Report on Robotics Ethics* (UNESCO: Pariz), <https://unesdoc.unesco.org/ark:/48223/pf0000253952> [stranica posjećena: 8. svibnja 2019.]

COMEST 2019. *Preliminary Study of the COMEST Extended Working Group on the Ethics of Artificial Intelligence* (UNESCO: Pariz), <https://unesdoc.unesco.org/ark:/48223/pf0000367823> [stranica posjećena: 8. svibnja 2019.]

Duhigg, C. 2012. *The Power of Habit: Why We Do What We Do in Life and Business* (New York: Random House).

Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, 2019. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, dostupno na: <https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html> [stranica posjećena: 8. svibnja 2019.]

Furnas, A. 2012. "It's not all about you: What privacy advocates don't get about data tracking on the web", *The Atlantic*, 15. ožujka; <https://www.theatlantic.com/technology/archive/2012/03/its-not-all-about-you-what-privacy-advocates-dont-get-about-data-tracking-on-the-web/254533/> [stranica posjećena: 8. svibnja 2019.]

Galliot, J. 2015. *Military Robots: Mapping the Moral Landscape* (Ashgate: Farnham).

Giubilini, A. i Savulescu, J. 2018. "The artificial moral advisor: The 'ideal observer' meets artificial intelligence", *Philosophy and Technology* 31(2): 169-188.

Gogoll, J. i Müller, J. F. 2017. "Autonomous cars: In favor of a mandatory ethics setting", *Science and Engineering Ethics* 23(3): 681-700.

Harris, J. 2007. *Enhancing Evolution: The Ethical Case for Making Better People* (Princeton University Press: Princeton / Oxford).

- Kant, I. 2016. [1785] *Utemeljenje metafizike čudoređa*, prev. J. Talanga (KruZak: Zagreb).
- Krishnan, A. 2009. *Killer Robots: Legality and Ethicality of Autonomous Weapons* (Ashgate: Farnham)
- Lokhorst, G.-J. i van den Hoven, J. 2012. "Responsibility for military robots", u: P. Lin, K. Abney i G. Bekey (ur.), *Robot Ethics: The Ethical and Social Implications of Robotics* (MIT Press: Cambridge, Mass.), 145-156.
- Lu, J. 2016. "Will medical technology deskill doctors?", *International Education Studies* 9(7): 130-134.
- McCarthy, J., Minsky, M. L., Rochester, N. i Shannon C. E. 2006. [1955], "A proposal for the Dartmouth summer research project on artificial intelligence", *AI Magazine* 27(4): 12-14.
- Mill, J. S. 1998. [1863] *Utilitarianism* (Oxford University Press: Oxford).
- Millar, J. 2017. "Ethics settings for autonomous vehicles", u: P. Lin, R. Jenkins i K. Abney (ur.), *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence* (Oxford University Press: New York), 20-34.
- Müller, M. U. 2018. "Medical applications expose current limits of AI", *Spiegel Online*, 3. kolovoza, <https://www.spiegel.de/international/world/playing-doctor-with-watson-medical-applications-expose-current-limits-of-ai-a-1221543.html> [stranica posjećena: 8. svibnja 2019.]
- Nozick, R. 2003. [1974] *Anarhija, država i utopija*, prev. B. Jakovlev (Naklada Jesenski i Turk: Zagreb).
- Nyholm, S. 2018a. "The ethics of crashes with self-driving cars: A roadmap (I-II)", *Philosophy Compass* 13(7).
- Perkowitz, S. 2004. *Digital People: From Bionic Humans to Androids* (Joseph Henry Press: Washington).
- Primorac, I. 2006. *Etika na djelu: Ogledi iz primijenjene etike* (KruZak: Zagreb).
- Rawls, J. 1999. [1971] *A Theory of Justice. Revised Edition* (Harvard University Press: Cambridge).
- Reese, B. 2018. *The Fourth Age: Smart Robots, Conscious Computers, and the Future of Humanity* (Atria International: New York).
- Rosenberg, J. M. 1986. *A Dictionary of Artificial Intelligence and Robotics* (John Wiley & Sons: New York).
- Ross, W. D. 2002. [1930] *The Right and the Good* (Oxford University Press: Oxford).
- Russell, S. J. i Norvig, P. 2016. *Artificial Intelligence: A Modern Approach* (Pearson: Harlow).
- Sandbu, M. 2017. "Technological justice", *Financial Times*, 20. veljače 2017.
- Sandel, M. 2009. "The case against perfection: What's wrong with designer children, bionic athletes, and genetic engineering", u: J. Savulescu i Nick Bostrom (ur.), *Human Enhancement* (Oxford University Press: Oxford), 71-89.

- Scheutz, M. 2012. "The inherent dangers of unidirectional bonds between humans and social robots", u: P. Lin, K. Abney i G. Bekey (ur.), *Robot Ethics: The Ethical and Social Implications of Robotics* (MIT Press: Cambridge, Mass.), 205-221.
- Searle, J. 1980. "Minds, brains, and programs", *Behavioral and Brain Sciences* 3: 417-457.
- Sharkey, N. 2012. "Killing made easy: From joysticks to politics", u: P. Lin, K. Abney i G. Bekey (ur.), *Robot Ethics: The Ethical and Social Implications of Robotics* (MIT Press: Cambridge, Mass.), 111-128.
- Sharkey, N. i Sharkey, A. 2012a. "Robotic surgery and ethical challenges", u: P. Gomes (ur.), *Medical Robotics: Minimally Invasive Surgery* (Woodhead Publishing: Oxford), 276-291.
- Sharkey, N. i Sharkey, A. 2012b. "The rights and wrongs of robot care", u: P. Lin, K. Abney i G. Bekey (ur.), *Robot Ethics: The Ethical and Social Implications of Robotics* (MIT Press: Cambridge, Mass.), 267-282.
- Siegel, E. 2013. *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die* (Wiley: Hoboken).
- Singer, P. 2003. [1979] *Praktična etika*, prev. T. Bracanović (KruZak: Zagreb).
- Sparrow, R. 2007. "Killer robots", *Journal of Applied Philosophy* 24(1): 62-77.
- Strawser, B. J. 2013. (ur.) *Killing by Remote Control: The Ethics of Unmanned Military* (Oxford University Press: Oxford / New York).
- Talanga, J. 1999. *Uvod u etiku* (Hrvatski studiji: Zagreb).
- Turing, A. M. 1950. "Computing machinery and intelligence", *Mind* 59(236): 433-460.
- Turkle, S. 2012. *Sami zajedno: Zašto očekujemo više od tehnologije, a manje jedni od drugih*, prev. G. Blažanović (TIM press: Zagreb).
- van de Poel, I. i Royakkers, L. 2011. *Ethics, Technology and Engineering: An Introduction* (Wiley-Blackwell: Oxford).

Tomislav Bracanović

Institut of Philosophy, Zagreb

tomislav@ifzg.hr

THE ETHICAL CHALLENGES OF ARTIFICIAL INTELLIGENCE AND ROBOTICS

Abstract: The paper addresses several ethical challenges of artificial intelligence and robotics. Following the introductory remarks about ethical aspects of engineering and an account of three basic theories of normative ethics, a threat to the freedom of will, meaningfulness of ethics and human specialness is considered as an implicit challenge. The central part consists of an analysis of seven explicit challenges: autonomous vehicles, autonomous weapons systems, social robotics, the medical use of artificial intelligence and robotics, predictive analytics, the influence of artificial intelligence and robotics on human employment, and application of new technologies in human enhancement. The conclusion proposes some guidelines as to the way the ethics of artificial intelligence and robotics should be done, as well as to their possible influence on the way ethics itself is done and understood.

Keywords: Artificial intelligence, robotics, ethics